2-day Sub-Module – Introduction to Sequence Storage Formats

Session	Topic
1	Next-Generation Sequencing (NGS) & Sequence Reads Formats
2	Mapping Formats
3	Variant Formats
4	Nucleotide Formats
5	Protein Formats
6	Alignment & Tree Formats

Course Description:

This basic level bioinformatics sub-module provides participants with the opportunity to gain hands on experience working with numerous sequence storage formats and their inter-conversions. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform the required tasks. Course material includes discussions and exercises on sequence read formats, mapping formats, variant formats, nucleotide and protein formats and alignment and tree formats. The discussions will focus on the description of the format, type of data stored, common software that can read the format and conversion of one format into other equivalent formats. Bioinformatics programs for which instruction is provided include graphical user interfaces (GUIs), webservers and command line tools such as FastQC, Mauve, IGV, Plink, BWA, SAMtools, Jmol, Cytoscape, MEGA, NCBI and EBI among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in understanding the content, types and application of sequence storage formats. More specifically, participants will gain demonstrable understanding in the following areas:

- Type of information contained within different sequence storage formats
- Sequence format inter-conversions
- Knowledge of the correct use for various sequence storage formats

2-day Sub-Module – Sequence Alignments

Session	Topic
1	Pairwise Sequence Alignment
2	Multiple Sequence Alignment
3	Basic Local Alignment & Search Tool
4	Genome Mapping
5	Genome Alignment I
6	Genome Alignment II

Course Description:

This basic level bioinformatics sub-module provides participants with the opportunity to gain hands on experience working with different methods used for the alignment of DNA and protein sequences. Course material includes discussions and exercises on the understanding of tools and methods to align a pair of or multiple short sequences, utilization of sequence alignment for database searching, alignment of short read sequences to a genome and pairwise of multiple genome alignment. The scope is restricted to sequence alignments and related topics will be touched upon briefly. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform the required tasks. Bioinformatics programs for which instruction will be provided include graphical user interfaces (GUIs), webservers and command line tools such as the EMBOSS suite, MEGA, BLAST, BWA, CCT and Mauve among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in correctly performing sequence alignments and interpreting the results. More specifically, participants will gain demonstrable understanding in the following areas:

- Performing pairwise and multiple sequence alignments on small sequences
- Alignment application: BLAST and its variants
- Mapping of sequencing reads to a genome
- Generating, visualizing and interpreting genome alignments

2-day Sub-Module – Bioinformatics Programming Basics

Session	Topic
1	Introduction to Linux Environment
2	Linux File System & Basic System Administration
3	Utility Compilation and Installation with *nix
4	Basic File Handling
5	Accessing and retrieving files from Linux
6	Wrap Up

Course Description:

This basic level bioinformatics sub-module provides participants with the essential skills required to be able to work in a Linux/*nix system when performing bioinformatics analyses. The participants will receive instruction in basic command line Linux and system administration and will be instructed in the compilation and installation of software in the Linux environment. Topics covered include the basics of Linux and file systems, file access, file handling, pipeline development and shell scripting. At the conclusion of the course, participants will have sufficient experience with Linux to begin using it in their research. The whole course will be geared towards the basic application of Linux in bioinformatics.

SMART Learning Objective:

By the end of the course, the participants are expected to have proficiency in utilizing Linux's command line interface. More specifically, participants will gain demonstrable understanding in the following areas:

- File handling and manipulation and program installation in a Linux/*nix environment
- Accessing and retrieving data from the web
- Pipeline development & shell scripting

2-day Sub-Module – Introduction to Bioinformatic Databases

Session	Topic
1	Bioinformatics Background & Sequence Formats
2	Basic Local Alignment & Search Tool (BLAST)
3	NCBI
4	Functional Annotation Databases
5	Multi-Locus Sequence Typing (MLST)
6	Custom Genome Databases

Course Description:

This basic level bioinformatics sub-module provides participants with an overview of different database resources available for bioinformatics along with some discussions on the implementation of bioinformatics databases and their associated tools. Course material includes an introduction to bioinformatics, next-generation sequencing (NGS) technologies, database searching utilities and the basics of setting up custom built databases. Databases that will be discussed include NCBI, specialized databases for functional annotation and PubMLST. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analyses.

SMART Learning Objective:

By the end of the course participants are expected to have understanding in utilizing bioinformatics databases. More specifically, participants will gain demonstrable understanding in the following areas:

- Understanding differences between primary and secondary databases
- Sequence storage formats and conversions
- Applications of databases in specific areas of research

2-day Sub-Module – Computational Phylogenetics

Session	Topic
1	Pairwise Sequence Alignment I
2	Pairwise Sequence Alignment II
3	Multiple Sequence Alignment
4	Phylogenetic Reconstruction I
5	Phylogenetic Reconstruction II
6	Genome Level Comparisons

Course Description:

This basic level bioinformatics sub-module provides participants methods for aligning sequences and reconstructing phylogenies as a way to discover relationships between organisms through use of molecular data. Course material includes the basics of sequence alignment and the corresponding assumptions, pairwise sequence alignment, multiple sequence alignment, basics of phylogenetics and interpreting trees, methods of tree construction and their comparisons, tree view and editing and distance matrix computation. The methods covered focus on small length sequences, not long sequences such as genomes where different methods are required. Tools that are used include the EMBOSS suite, MEGA and JSpecies among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in correctly reconstructing and interpret phylogenies. More specifically, the participants will gain demonstrable understanding in the following areas:

- Aligning sequences for phylogeny reconstruction
- Reconstructing and interpreting phylogenies
- Quantitative genome level comparisons

2-day Sub-Module – MLST Based Molecular Epidemiology

Session	Topic
1	Phylogenetic Basics
2	Multi-Locus Sequence Typing (MLST)
3	Basic Local Alignment & Search Tool (BLAST)
4	Next-Generation Sequencing (NGS)
5	Genome Assembly & Mapping
6	Variant Calling

Course Description:

This intermediate level bioinformatics sub-module provides participants the concepts and methods for performing multi-locus sequence typing (MLST) for pathogen identification. Course material includes discussions and exercises on the evolutionary underpinning of MLST based typed methods, the concept of MLST, its current implementation, associated resources and tools, the basics of next-generation sequencing (NGS) technologies and their influence on MLST and variant detection for SNP based typing. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis. A mix of graphical user interface (GUIs), webservers and command line tools are included such as PubMLST, MEGA, BLAST, BWA, SAMtools, BCFtools and IGV among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in performing genomic based multi-locus sequence typing. More specifically, participants will gain demonstrable understanding in the following areas:

- MLST from allele sequences
- MLST from whole genomes
- Understanding the nuances of MLST
- Detecting SNPs for SNP-based typing schemes

2-day Sub-Module – Introduction to NCBI

Session	Topic
1	File Formats
2	NCBI
3	Introduction to Scientific Computing
4	Sequence Based Database Searching
5	Programmatic Access to NCBI
6	NCBI Genome Workbench

Course Description:

This intermediate level bioinformatics sub-module provides participants with an understanding of the National Center for Biotechnology Information (NCBI), a pivotal resource in facilitating bioinformatics research by enabling data storage, retrieval and analysis. Course material covers discussions on the structure of NCBI, different component databases and their inter-connections, numerous analytical tools available in NCBI, Entrez, and the programmatic and web-based query and cess of data stored in NCBI databases. Bioinformatics programs for which instruction is provided include webservers and command line tools such as NCBI Genome Workbench, BLAST, E-Direct and E-Utils among others. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis however familiarity on Linux will be beneficial.

SMART Learning Objective:

By the end of the course, the participants are expected to have proficiency in using the NCBI databases and its associated tools. More specifically, participants will gain demonstrable understanding in the following areas:

- Access and retrieval of data from NCBI and its components
- Understanding different analytical tools and their applications
- Learning how to programmatically connect to NCBI and access data

2-day Sub-Module – Protein Functional Annotation

Session	Topic
1	Bioinformatics Background & Sequence Formats
2	Basic Local Alignment & Search Tool (BLAST)
3	Functional Annotation Databases
4	NCBI
5	Pathway Analysis
6	Clustering Analysis

Course Description:

This intermediate level bioinformatics sub-module provides participants with the methods and techniques for functionally annotating proteins and non-coding RNAs as a way to attach biological information to genomic elements. Course material includes an introduction to bioinformatics and sequence formats, annotation via homology, *ab initio* annotation, accessing data from NCBI, clustering analysis, using KEGG and other functional annotation databases and associated tools. Exercises consist of performing functional annotations both via webservers and the Linux command-line. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis. Bioinformatics programs for which instruction will be provided include webservers such as Pfam, InterPro and RAST and command line tools such as BLAST+ and BLAST among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in the annotation of proteins and other genomic elements. More specifically, participants will gain demonstrable understanding in the following areas:

- Homology based annotation
- *Ab initio* based annotation
- Protein clustering analysis
- Functional annotation databases and associated tools

2-day Sub-Module – Intermediate Bioinformatics Programming

Session	Topic
1	Introduction to Programming Concepts and Perl
2	Sequences and Data Structure
3	More Data Manipulation Operations
4	Loops and Control
5	Files and Devices
6	Subroutines and Modules

Course Description:

This intermediate level bioinformatics sub-module provides participants with the opportunity to gain hands on experience working with the Perl programming language, a commonly used tool in the analysis of biological data. Course material includes the basics of programming concepts and Perl, file access, file handling and manipulation, loops, subroutines and modules. By the end of the course, participants will have sufficient knowledge to develop their own working Perl scripts and pipelines that can be applied to their research. The whole course will be geared towards application of Perl in bioinformatics.

SMART Learning Objective:

By the end of the course, the participants are expected to have demonstrable skill in Perl and creating their own pipelines and scripts. More specifically, participants will have understanding in the following areas of Perl and bioinformatics:

- Programming paradigms
- Basic data structures and related operations
- File handling and manipulation
- Scripts and pipeline development

2-day Sub-Module – Sequence Visualization

Session	Topic
1	Multiple Sequence Alignment & Sequence Logos
2	Trees – distance & character based
3	Genome Mapping
4	Genome Alignment & Visualization
5	Circular Genome Views
6	Network Visualization

Course Description:

This intermediate level bioinformatics sub-module provides participants with the opportunity to gain hands on experience with sequence visualization. Course material includes discussions and exercises on the visualization of different types of sequences and analyses such as multiple sequence alignments, trees, read to genome mapping, genome alignments, circular genome visualization and network visualization. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis. Bioinformatics programs for which instruction is provided include graphical user interface (GUI) and command line tools such as MEGA, Mauve, CGView Comparison Tool (CCT), BWA, SAMtools and Cytoscape among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in successfully visualizing different types of sequences. More specifically, participants will gain demonstrable understanding in the following areas:

- Generating and visualizing multiple sequence alignments and associated sequence logos
- Read to genome mapping and visualization
- Genome alignment and comparison visualizations
- Protein/gene network visualizations

2-day Sub-Module – Gene Prediction

Session	Topic
1	Introduction to Scientific Computing
2	Prokaryotic Gene Prediction
3	Eukaryotic Gene Prediction
4	Sequence Similarity Searching
5	Hands On Session
6	Metagenomics Gene Prediction

Course Description:

This advanced level bioinformatics sub-module provides participants with the opportunity to gain hands on experience with the techniques and methods of predicting genes and features from complete genomes and assemblies using *ab initio* and homology based methods. Course material includes discussions and exercises on the prediction of genes/features in both prokaryotic and eukaryotic organisms as well as in metagenomics samples, focusing on why and how the techniques for prediction differ. Also included are discussion on the algorithms and paradigms behind gene prediction and the concept of merging prediction sets to generate a more complete set. Exercises consist of performing gene (and non-coding RNA) predictions and then merging the predictions to result in a final complete set. The participants will receive instruction in Linux command line and gain experience working at the command line to perform bioinformatics analysis. Bioinformatics programs for which instruction is provided include webservers and command line tools such as GeneMark Suite, Glimmer, Prodigal, RNAmmer, AUGUSTUS and BLAST among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in the prediction of prokaryotic and eukaryotic genes and other genomic features. More specifically, participants will gain demonstrable understanding in the following areas:

- Differences between prokaryotic and eukaryotic prediction paradigms
- Homology based prediction methods
- *Ab initio* based prediction methods
- Non-coding RNA predictions

2-day Sub-Module – Comparative Genomics

Session	Topic
1	Multiple Sequence Alignment
2	Sequence Based Database Searching
3	Average Nucleotide Identity (ANI)
4	Genome Alignment
5	Clustering Analysis
6	Wrap Up

Course Description:

This advanced level bioinformatics sub-module provides participants with methods of comparisons of different genomic features through sequence comparisons. Course material includes discussions and exercises on the techniques for aligning molecular sequences and methods for performing genome-scale comparisons. The participants will be instructed on aligning multiple short sequences, sequence based database searching, aligning a pair of or multiple genomes and gene/protein clustering. Bioinformatics programs for which instruction will be provided include graphical user interface (GUI), webservers and command line tools such as MEGA, BLAST suite, JSpecies and Mauve among others. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analyses.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in the performance of bioinformatics comparative analyses. More specifically, participants will gain demonstrable understanding in the following areas:

- Sequence alignments and visualizations
- Gene/protein level quantitative comparisons
- Genome level quantitative comparisons

2-day Sub-Module – Genome Assembly I

Session	Topic
1	Next-Generation Sequencing (NGS)
2	Introduction to Scientific Computing
3	Quality Control & Assessment
4	Illumina Sequencing (Paired end)
5	Assembly Comparison & Merging
6	Wrap Up

Course Description:

This advanced level bioinformatics sub-module provides participants with the opportunity to gain hands on experience in handling and assembly of next-generation sequencing (NGS) data. Course material includes bioinformatics aspects of sequencing basics, Illumina paired end sequencing, and *de novo* genome assembly. The module will involve a mix of graphical user interface (GUI) and command line tools such as FastQC, PRINSEQ, Velvet, SOAPdenovo2, QUAST and CISA among others. The participants will receive instruction in command line Linux and gain experience working at the command line to perform bioinformatics analysis of NGS data.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in correctly performing genome assemblies. More specifically, participants will gain demonstrable understanding in the following areas:

- Quality assessment and control of sequencing reads
- Assembly of reads from Illumina pair end platforms
- Assembly comparison and merging

2-day Sub-Module – Genome Assembly II

Session	Topic
1	Next-Generation Sequencing (NGS)
2	Introduction to Scientific Computing
3	Quality Control & Assessment
4	Genome Mapping
5	SNP Calling
6	Metagenomics assembly

Course Description:

This advanced level bioinformatics sub-module provides participants with the opportunity to gain hands on experience in the handling and analysis of next-generation sequencing (NGS) data. Course material includes bioinformatics aspects of sequencing basics, genome mapping and variant calling. Bioinformatics programs for which instruction will be provided include a mix of graphical user interface (GUI) and command line tools such as FastQC, PRINSEQ, BWA, SAMtools, BCFtools and IGV among others. The participants will receive instruction in command line Linux and gain experience working at the command line to perform bioinformatics analysis of NGS data.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in correctly performing read to genome mapping and visualization. More specifically, participants will gain demonstrable understanding in the following areas:

- Quality assessment and control of sequencing reads
- Genome mapping and SNP calling